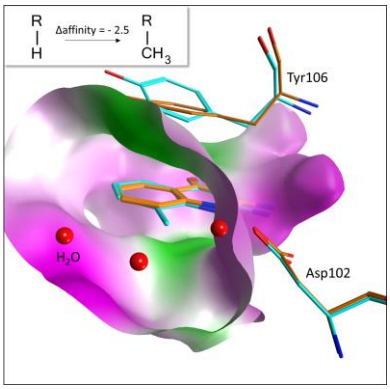
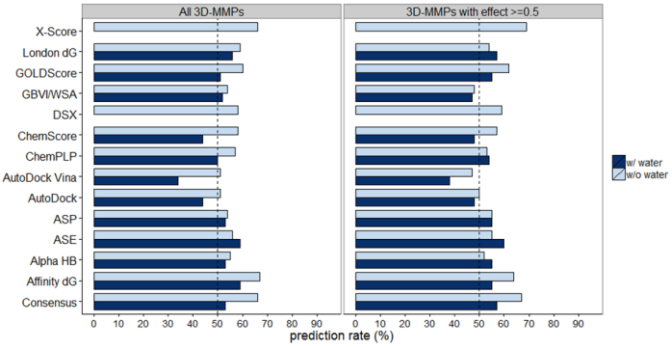
A Diverse Test Set for the Validation of Scoring Functions based on Matched Molecular Pairs

Lena Kalinowsky[1,†], Julia Weber[1,†], Ewgenij Proschak[*,1].

1. Institute of Pharmaceutical Chemistry, Goethe-University of Frankfurt, Max-von-Laue Str. 9, Frankfurt a.M., D-60438 Germany.

†. L.K. and J.W. contributed equally and share first authorship





In structure-based drug design the prediction of protein-ligand interactions and their contribution to the binding free energy is a challenging task. Scoring function evaluation has shown that docking already gives valuable results. However, the "scoring" problem is still a very ambiguous. Today, scoring functions are not able to precisely predict the binding free energy of protein-ligand complexes. In this study we established a diverse data set of 99 Matched Molecular Pairs (3D-MMPs). This data set was used to study the predictive power of scoring functions and to investigate their disadvantages. The 13 most commonly used scoring functions (i.a. MOE, GOLD, AutoDock 4.2) have been used to score and evaluate the binding free energy predictive capability. None of the scoring functions reached a satisfactory result in our evaluation. Only two scoring functions reached a prediction rate of more than 60% in the prediction of the trend of a transformation effect. By analyzing the correlation between the score and the molecule size we could show that in 67% the affinity increases when the size of the molecules increases. Most of the scoring functions themselves correlate more with the changes in molecule size than with the changes in binding affinity.